MSR56

Incorporating background mortality into survival extrapolations: determining the accuracy using a simulation study

Hordijk DD[1], Dunnewind N[1], Gittfried AMV[1], Pouwels X[2], Koffijberg H[2]

[1]OPEN Health HEOR & Market Access, Rotterdam, The Netherlands,
[2]University of Twente, Enschede, The Netherlands

OPEN HEALTH

INTRODUCTION

- Clinical trial data is often limited in terms of follow-up time while HTA submissions frequently require assessment of long-term survival outcomes, thus survival extrapolations are necessary.
- Recent NICE guidance[1] suggests employing a relative survival approach to incorporate general population background mortality (GPM) into survival extrapolations, for example to adjust for the increasing GPM as patients age beyond follow-up.
- Alternative methodologies to incorporate GPM within survival extrapolations are also available, yet guidance on the appropriateness of these methods across situations is lacking, and no selection algorithm exists on which method to use.[1,2]

OBJECTIVES

- To perform a simulation study and compare the performance of different methods for incorporating GPM information into survival extrapolations across a large set of scenarios to identify patterns in performance between the different methods to improve survival extrapolation guidance.

METHODS

- A simulation was implemented in R[3] that could simulate patient datasets for 324 different scenarios to assess the performance of GPM incorporating methods by comparing their extrapolations to a known overall survival.
- Various characteristics were generated per patient based on the scenario. Then, GPM and disease-specific mortality (DSM) event times were generated based on the patient characteristics and scenario. The lowest time was used for the all-cause mortality (ACM) event time.
- The datasets were artificially censored by drawing a censoring time from an exponential distribution with rate $\lambda_{censoring}$, and setting the event (or censoring) time to the lowest out of the ACM time, censoring time and the follow-up length. Known survival was obtained using the uncensored event times.
- Performance was compared in terms of mean absolute error on extrapolated overall mean survival, survival probability at time t (t = follow-up * 2.5) and restricted mean survival time (RMST) until the end of follow-up.
- 5,000 replications were performed for each scenario.

Scenarios

- Scenarios were defined by varying the age, distribution and level of DSM survival, heterogeneity and the level of information. See Table 1 for a complete overview.
- To obtain survival models for generating DSM event times that reflect differing levels of survival and have differently shaped survival curves without the need for manually selecting parameters, parametric models were fit to three real-world datasets using four parametric distributions.

Table 1. Overview of scenarios for simulation

Dimension *(Level)*	Definition	# of levels
Age	**N ~ (μ, σ_{age})**	**x 3**
Young, Average, Old	μ=35, μ=50, μ=75	
DSM survival	**Various survival levels**	**x 3**
Low	Pancreatic cancer (median survival of 17 months)[4]	*Follow-up = 1.2 years*
Medium	Myocarditis (5-year survival of 56%)[5]	*Follow-up = 6 years*
High	Ulcerative colitis (40-year survival of 59%)[6]	*Follow-up = 8 years*
Parametric distribution of DSM survival	Weibull, Gompertz, Log-logistic, Lognormal	**x 4**
Heterogeneity	**σ_{age}, division over sexes** *(either M/F or F/M)*	**x 3**
Low	σ_{age} = 3, 90% / 10%	
Medium	σ_{age} = 6, 70% / 30%	
High	σ_{age} = 12, 50% / 50%	
Level of information	**$n_{patients}$, $\lambda_{censoring}$, extrapolation knowledge**	**x 3**
Low	n=100, λ=0.3, summary	
Medium	n=300, λ=0.067, summary	
High	n=500, λ=0.01, full	

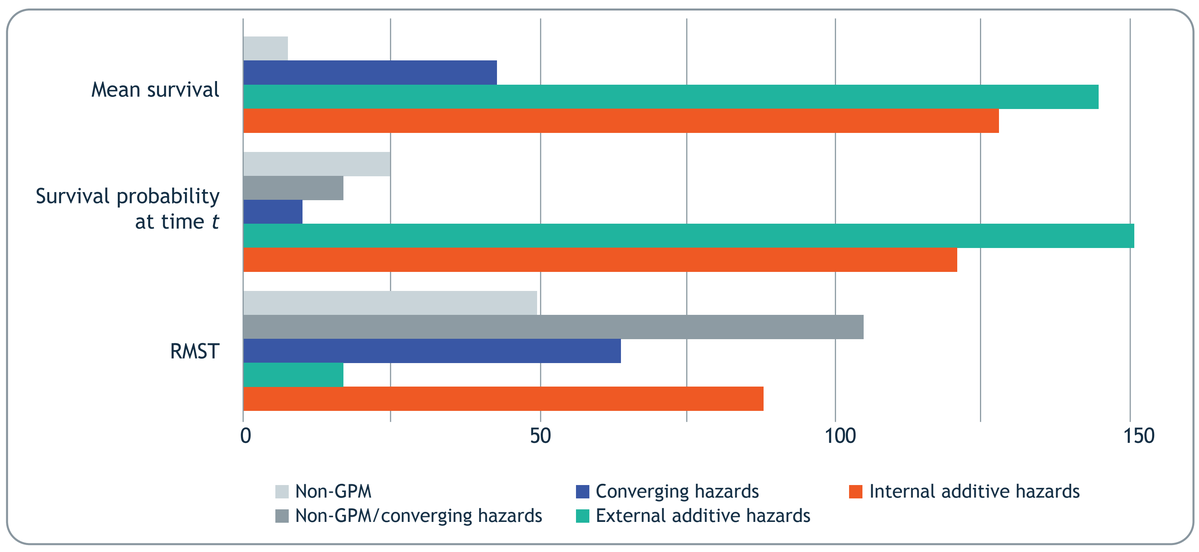
Extrapolation & GPM incorporating methods

- Three methods that incorporate GPM information into parametric survival extrapolations selected based on a previous study[7] were compared to non-GPM extrapolations. The GPM incorporating methods were as follows:
  - Internal additive hazards: using a relative survival approach[1], i.e., include GPM hazards within the log-likelihood function to derive DSM hazards, then add the DSM (or excess) hazards to GPM hazards for extrapolation.
  - External additive hazards: fit a parametric model to trial data to obtain ACM hazards, add GPM hazards separately for extrapolation.
  - Converging hazards: use the highest hazard out of the ACM hazard fit to the trial data (from the same parametric model used for the external additive hazards and non-GPM methods) and GPM hazard at each point in time for extrapolation.
- To fit the survival models used for extrapolation, standard parametric models as recommended by NICE[8] (exponential, Weibull, log-normal, log-logistic, Gompertz, generalized gamma) and the generalized F distribution were fit using flexsurv.[9]
- If 'summary' extrapolation knowledge was available (based on the level of information scenario), only the mean age of patients and division over sexes were assumed to be known, and the weighted GPM hazards were used for extrapolation and fitting the relative survival model. When 'full' extrapolation knowledge was assumed to be available, individual GPM hazards, based on patients' age and sex, were used.

RESULTS

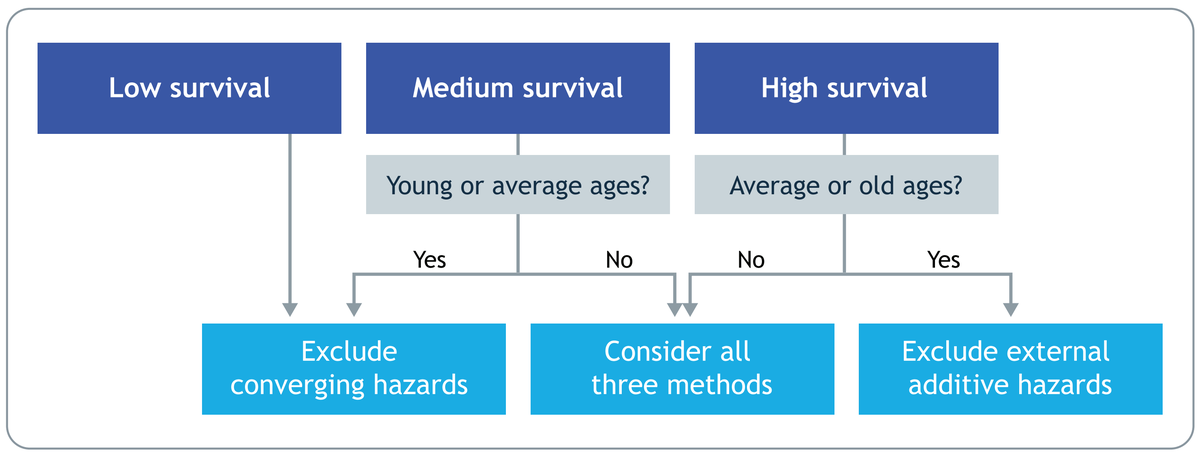
- Non-GPM extrapolations were outperformed (in terms of mean absolute error) in 256 scenarios (79% of 324) on average across the three estimands.
- Non-GPM extrapolations had the same mean absolute error as the converging hazards method in an average of 41 scenarios (13% of 324) and the lowest mean absolute error overall in 28 scenarios (9% of 324) on average across the three estimands.
- The best performing model in terms of mean absolute error did not always use the same parametric distribution as the parametric distribution used for generating the DSM event times.

Figure 1. Number of scenarios each GPM incorporating method had lowest mean absolute error for each estimand. Scenarios in which non-GPM and converging hazards extrapolations were identical are labelled 'Non-GPM/converging hazards'



- Large differences can be seen in Figure 1 between performance of the methods in estimating RMST and extrapolating mean survival and survival probability at time t.
- A consistent pattern to identify the best approach across the parametric distributions used for generating DSM survival times was not identified. However, certain methods consistently performed poorly when extrapolating survival in specific scenarios.
- Figure 2 presents a flowchart for selecting a GPM incorporating method, given the definitions shown in Table 1.

Figure 2. Flowchart for selecting a GPM incorporating method



CONCLUSIONS

- Methods that incorporate GPM information outperformed non-GPM extrapolations in a majority of scenarios across estimands overall. Furthermore, considering the low number of scenarios where non-GPM extrapolations performed best – especially when extrapolating overall mean survival - the study supports the current NICE guidance recommending to incorporate GPM information into survival extrapolations.
- A flowchart (Figure 2) was introduced, providing guidance on selecting a method for incorporating GPM into survival extrapolations in different scenarios.
- Real-world trials might not match the scenario definitions used in this study closely, reducing the usefulness of the flowchart presented in Figure 2. Thus, future studies could explore the use of continuous scales for scenario parameters to remove the need of arbitrary scenario definitions and provide more accurate guidance on which trial characteristics influence the performance of the methods incorporating GPM information.
- In future studies, performance of GPM incorporating methods on more complex relationships between GPM and DSM should be explored, for example by incorporating a cure fraction. This would also allow testing the impact of incorporating GPM information on more complex survival models.

REFERENCES

1. Rutherford, M., Lambert, P., Sweeting, M., Pennington, R., Crowther, M., Abrams, K., & Latimer, N. (2020). NICE DSU Technical Support Document 21. Flexible Methods for Survival Analysis.
2. Coyle, D., Haines, A., & Lee, K. (2023). Extrapolating Clinical Evidence Within Economic Evaluations. *Canadian Journal of Health Technologies*, 3(5).
3. R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
4. Kuhlmann, K. F. D., de Castro, S. M. M., Wesseling, J. G., ten Kate, F. J. W., Offerhaus, G. J. A., Busch, O. R.C., van Gulik, T. M., Obertop, H., & Gouma, D. J. (2004). Surgical treatment of pancreatic adenocarcinoma. *European Journal of Cancer*, 40(4), 549-558.
5. Magnani, J. W., Danik, H. J., Dec, G. W., Jr., & DiSalvo, T. G. (2006). Survival in biopsy-proven myocarditis: a long-term retrospective analysis of the histopathologic, clinical, and hemodynamic predictors. *Am Heart J*, 151(2), 463-470.
6. Jess, T., Loftus, E. V., Jr., Harmsen, W. S., Zinsmeister, A. R., Tremaine, W. J., Melton, L. J., 3rd, Munkholm, P., & Sandborn, W. J. (2006). Survival and cause specific mortality in patients with inflammatory bowel disease: a long term outcome study in Olmsted County, Minnesota, 1940-2004. *Gut*, 55(9), 1248-1254.
7. van Oostrum, I., Ouwens, M., Remiro-Azocar, A., Baio, G., Postma, M. J., Buskens, E., & Heeg, B. (2021). Comparison of Parametric Survival Extrapolation Approaches Incorporating General Population Mortality for Adequate Health Technology Assessment of New Oncology Drugs. *Value Health*, 24(9), 1294-1301.
8. Latimer, N. (2011). NICE DSU Technical Support Document 14: Undertaking survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data.
9. Jackson, C. (2016). "flexsurv: A Platform for Parametric Survival Modeling in R." *Journal of Statistical Software*, 70(8), 1–33.

Presented at ISPOR Europe 2023